Advancing High End Computing: Linking to National Goals

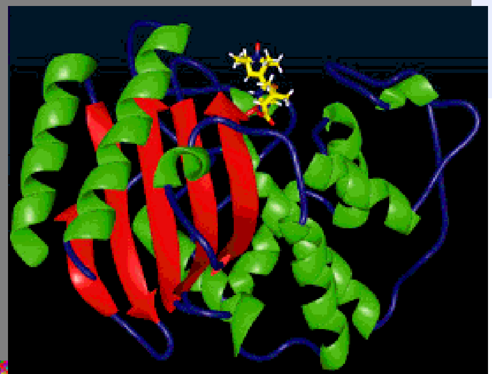
Comments by

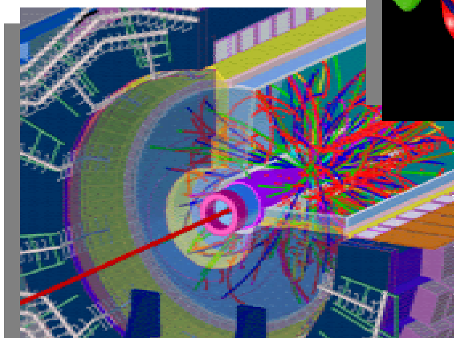
Horst D. Simon
Director, NERSC Center Division, LBNL
November 12, 2003
http://www.nersc.gov/~simon

Outline

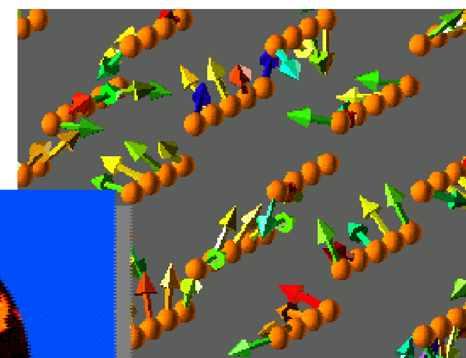
- Introducing NERSC
- The Evolution of HEC Systems and Applications: the Divergence Problem
- U.S. Leadership in HEC
- The Implementation of a Cyberinfrastructure

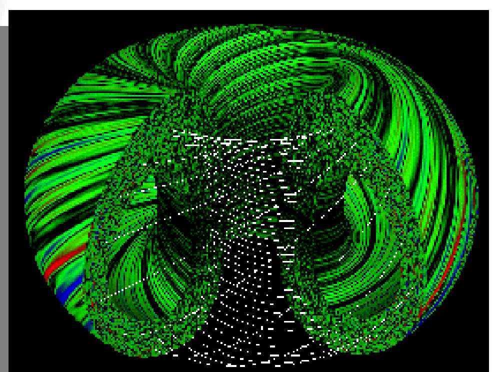
National Energy Research Scientific Computing Center

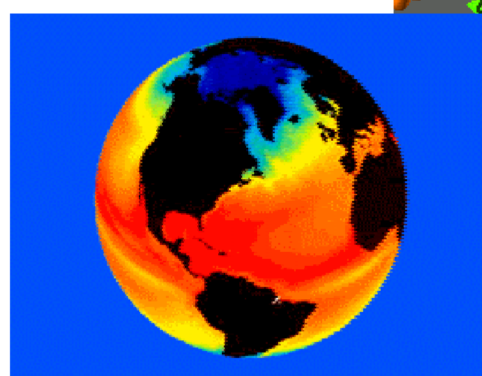




- Serves all disciplines of the DOE Office of Science
- ~2000 Users in ~400 projects
- Focus on large-scale computing







Upgraded NERSC 3 (Seaborg) Characteristics

- The upgraded “Seaborg” system has
  - 416 16-way IBM Power 3+ nodes with each CPU at 1.5 Gflop/s
    - 380 for computation
  - 6,656 CPUs – 6,080 for computation
  - Total Peak Performance of 10 Tflops
  - Total Aggregate Memory is 7.8 TB
  - Total GPFS disk will be 44 TB
    - Local system disk is an additional 15 TB
  - Combined SSP-2 is greater than 1.238 Tflop/s
  - NERSC 3E is in full production as of March 1,2003

21th List: The TOP10

Rank	Manufacturer	Computer	R_{max} [TF/s]	Installation Site	Country	Year	Area of Installation	# Proc
1	NEC	Earth-Simulator	35.86	Earth Simulator Center	Japan	2002	Research	5120
2	HP	ASCI Q, AlphaServer SC	13.88	Los Alamos National Laboratory	USA	2002	Research	8192
3	Linux Networx/ Quadrics	MCR Cluster	7.63	Lawrence Livermore National Laboratory	USA	2002	Research	2304
4	IBM	ASCI White SP Power3	7.3	Lawrence Livermore National Laboratory	USA	2000	Research	8192
5	IBM	Seaborg SP Power 3	7.3	NERSC Lawrence Berkeley Nat. Lab.	USA	2002	Research	6656
6	IBM/Quadrics	xSeries Cluster Xeon 2.4 GHz	6.59	Lawrence Livermore National Laboratory	USA	2003	Research	1920
7	Fujitsu	PRIMEPOWER HPC2500	5.41	National Aerospace Laboratory of Japan	Japan	2002	Research	2304
8	HP	rx2600 Itanium2 Cluster Qadrics	4.88	Pacific Northwest National Laboratory	USA	2003	Research	1536
9	HP	AlphaServer SC ES45 1 GHz	4.46	Pittsburgh Supercomputing Center	USA	2001	Academic	3016
10	HP	AlphaServer SC ES45 1 GHz	3.98	Commissariat a l'Energie Atomique (CEA)	France	2001	Research	2560

Outline

- Introducing NERSC
- The Evolution of HEC Systems and Applications: the Divergence Problem
- U.S. Leadership in HEC
- The Implementation of a Cyberinfrastructure

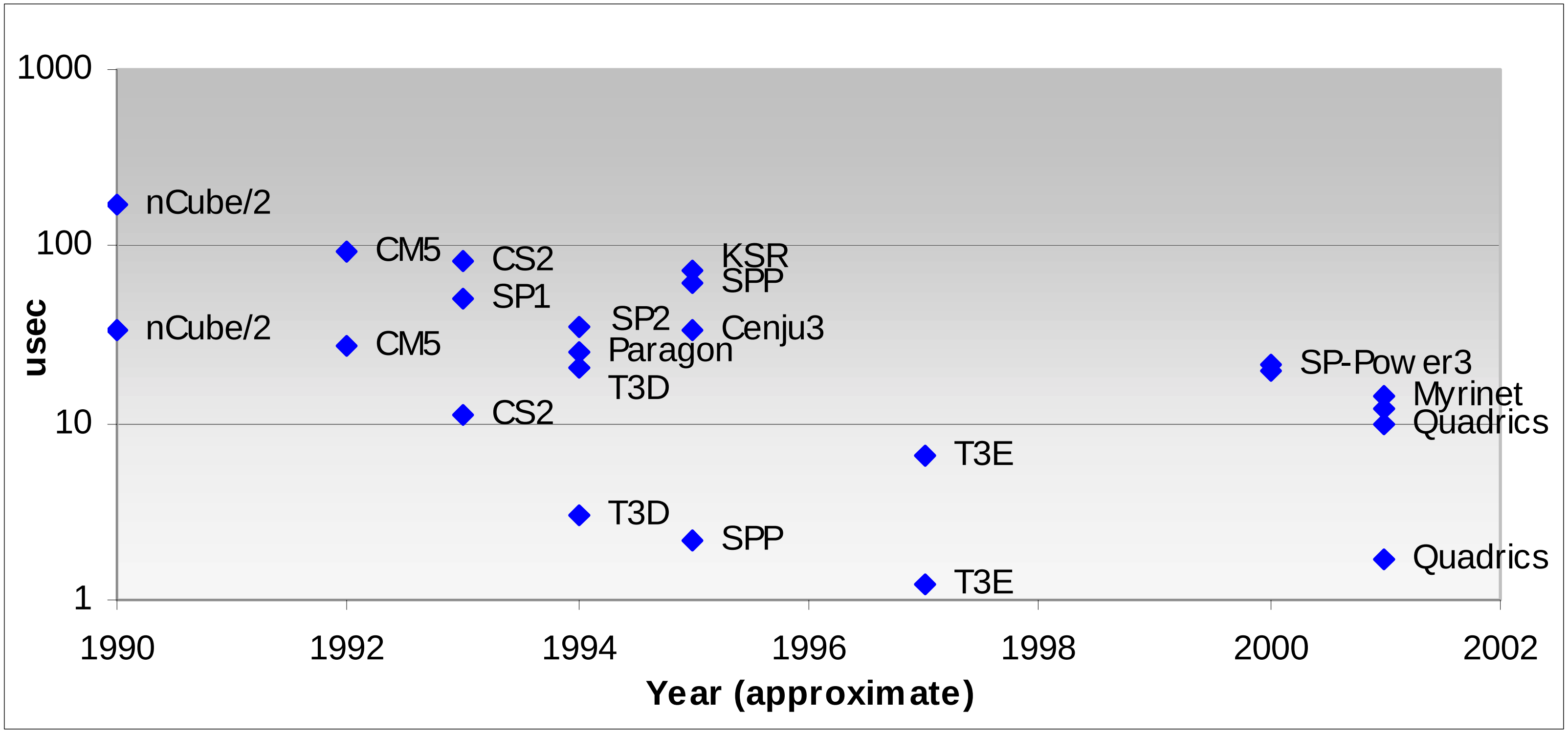
Rogers and Bozeman, pg. 14

- “ … during the 1990s … the gap in the development path of HEC computers and the needs of the user community diverged …”
- “… the government merely provided specifications and issued procurement requests, … leaving the market to its own forces …”

The State of the American Computer Industry – In Scientific Computing

- The major players that are still active in scientific supercomputing are
  - IBM
  - Hewlett Packard
  - Cray
  - Sun
  - SGI
- We don't have a building block optimized for scientific computation.
- The target commercial market is data and web serving, and that market dominates the economics of the computer industry beyond the personal computer.
- The architectural barriers for scientific computing stem from this situation
  - Memory bandwidth and latency (optimized for databases)
  - Interconnect bandwidth and latency (optimized for transaction processing)
- If you don't have a viable market for those building blocks, then how do you cause them to be created?

End to End Latency Over Time

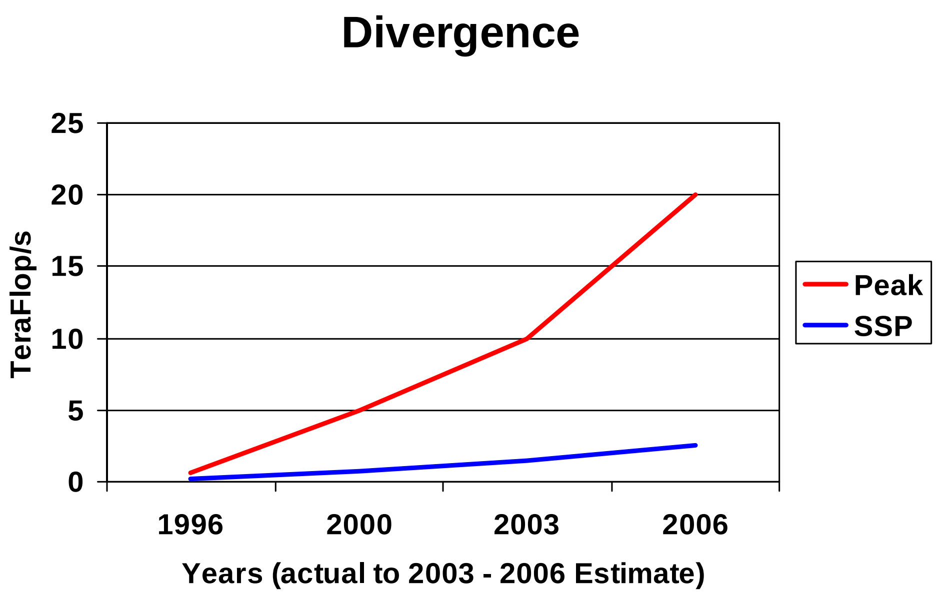


- **Latency has not improved significantly**
  - **T3E (shmem) was lowest point**
  - **Federation in 2003 will not reach that level – 7 years later!**

Data from C. Bell et al. “An Evaluation of Current High-Performance Networks” see http://upc.nersc.gov/publications

The Divergence Problem

- The requirements of high performance computing for science and engineering and the requirements of the commercial market are diverging.
- The commercial cluster of SMP approach is no longer sufficient to provide the highest level of performance
  - Lack of memory bandwidth
  - High interconnect latency
  - Lack of interconnect bandwidth
  - Lack of high performance parallel I/O
  - High cost of ownership for large scale systems



A Consensus

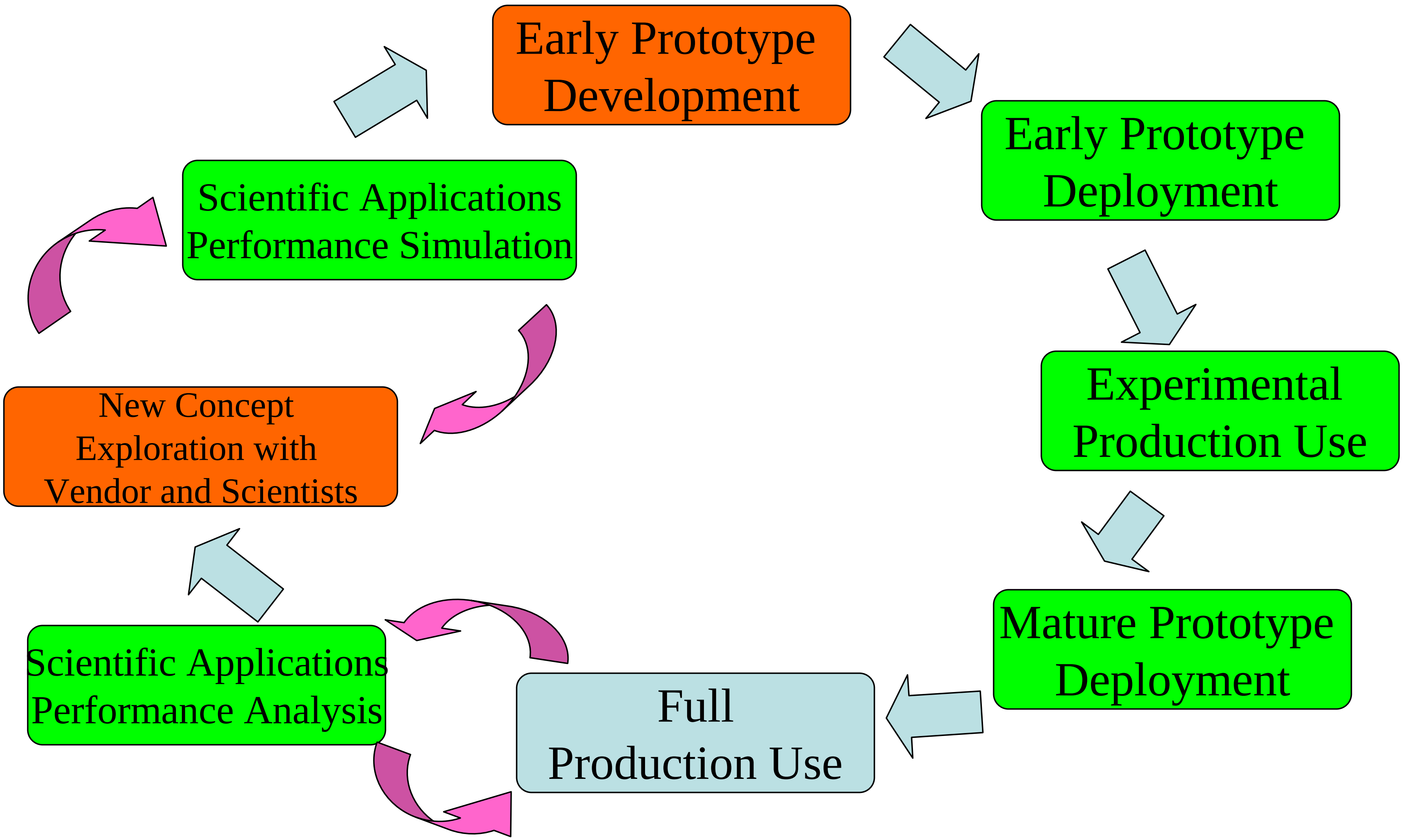
Seven DOE laboratories have co-authored a white paper to the High-End Computing Revitalization Task Force on:

"Creating Science-Driven Computer Architecture: A New Path to Scientific Leadership"

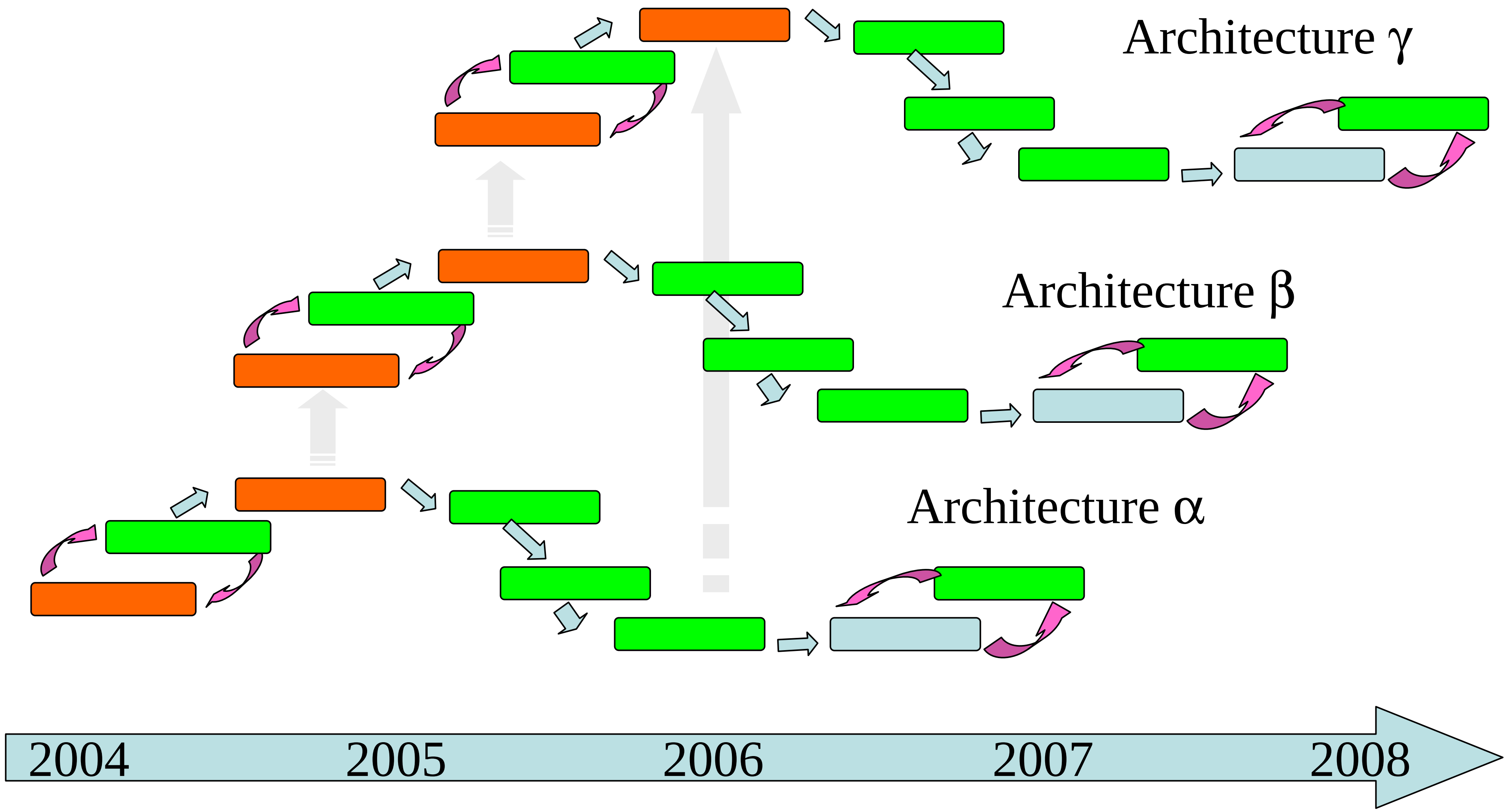
http://www.nersc.gov/~simon/Papers/HECRFT-V4-2003.pdf

- **Agreement on the problem**
  - **Communications and memory bandwidth are not scaling with peak processor power.**
  - **The commodity building block was the microprocessor, but is now the entire server (SMP).**
  - **Scientific computing has been driven to the logical extreme of COTS systems**
- **Agreement on the solution**
  - **Sustained cooperative development of new computer architectures**
  - **Focus on sustained performance of scientific applications**
  - **Strategy to pursue multiple science-driven architectures**

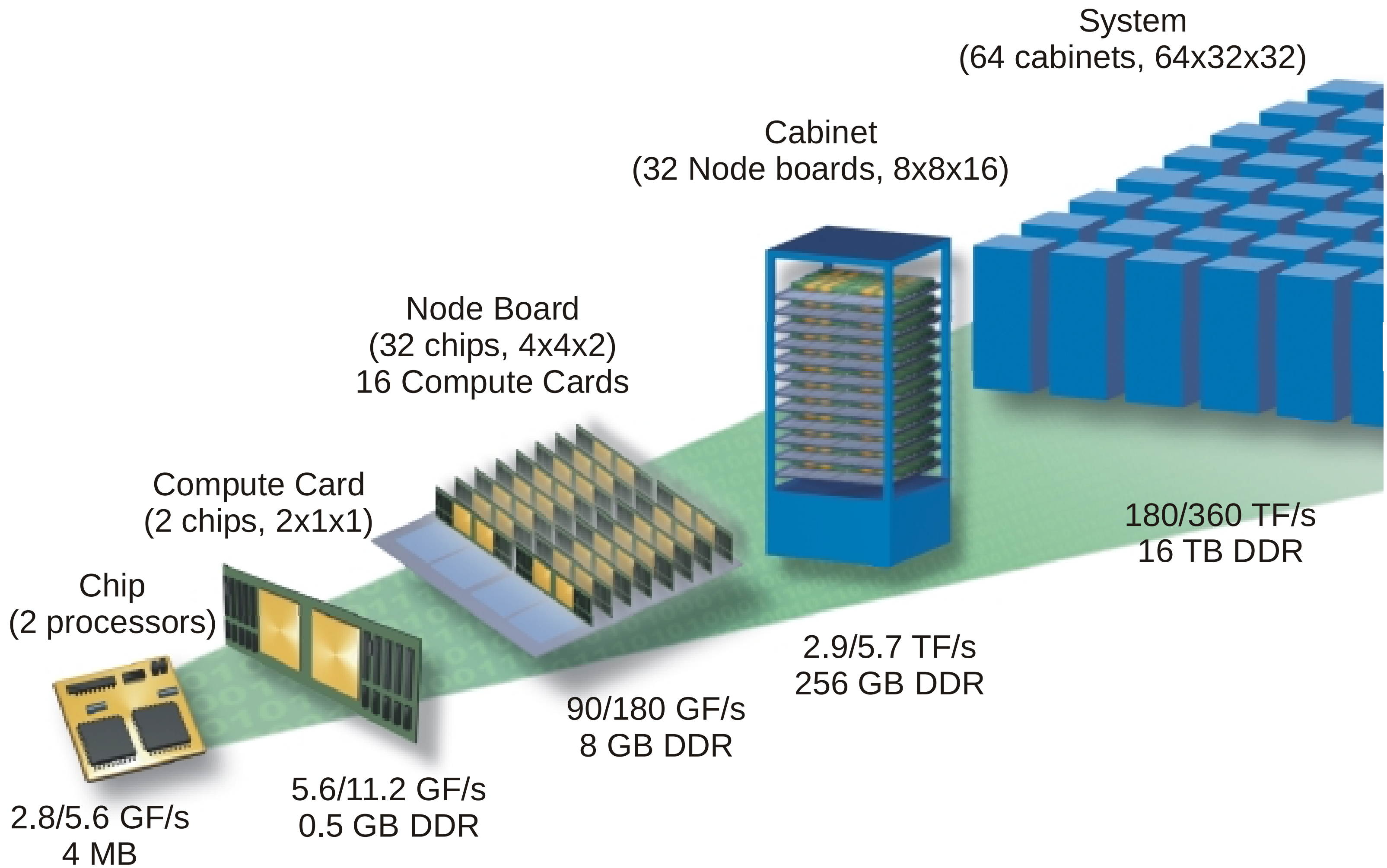
A Cycle of Design and Evaluation for a Single Architecture



A Sustainable Program of Multiple Cycles of Design, Evaluation and Scientific Production

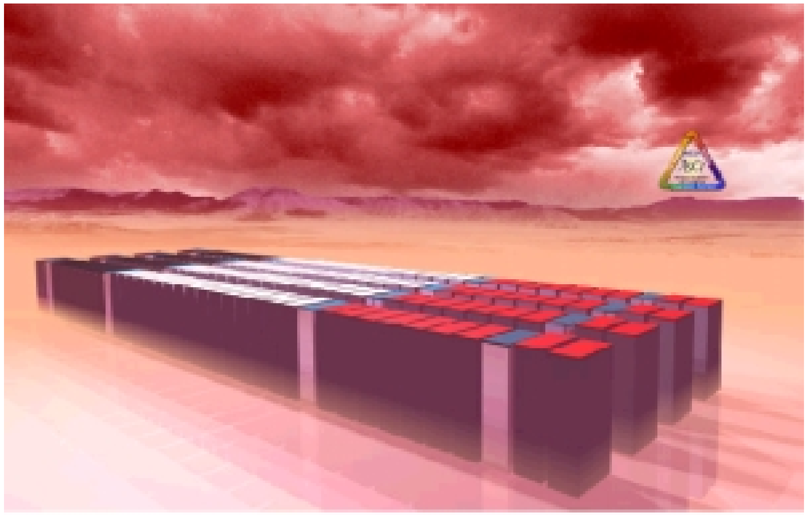


Example of Cooperative Development: Blue Gene/L (IBM and LLNL)



Example of Cooperative Development: Red Storm (Cray and Sandia Natl. Lab.)

- Collaboration between Sandia Natl. Lab. and Cray
- True MPP, designed to be a single system
- Distributed memory MIMD parallel supercomputer
- Fully connected 3-D mesh interconnect. Each compute node processor has a bi-directional connection to the primary communication network.
- 108 compute node cabinets and 10,368 compute node processors (AMD Sledgehammer @ 2.0 GHz) ~20 Tflop/s peak
- ~10 TB of DDR memory @ 333 MHz
- 240 TB of disk storage (120 TB per color)
- Less than 2 MW total power and cooling.
- Less than 3,000 square feet of floor space



Courtesy: Bill Camp and Jim Thompkins, Sandia

Outline

- Introducing NERSC
- The Evolution of HEC Systems and Applications: the Divergence Problem
- U.S. Leadership in HEC
- The Implementation of a Cyberinfrastructure

Rogers and Bozeman, pg. 21

- **Status Leadership: perceived to be the winner of a race**

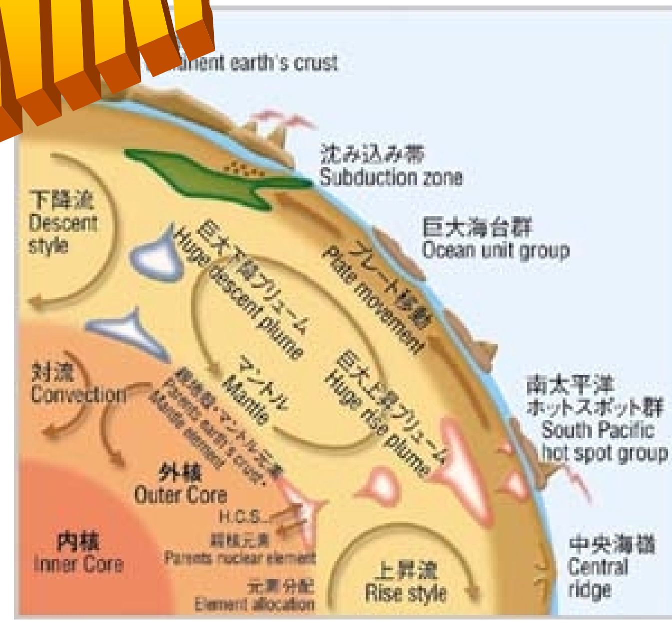
Versus

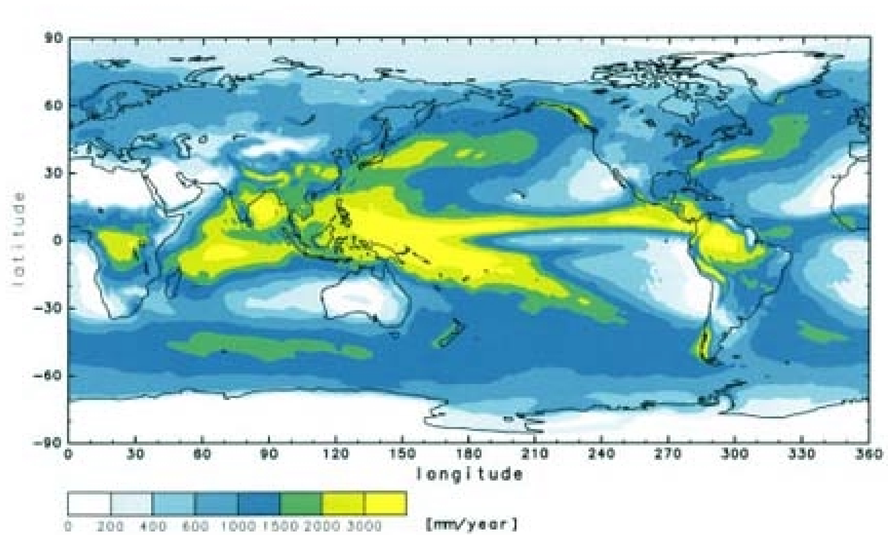
- **Pragmatic Leadership: ability to secure desired results into the future**
  - **National security leadership**
  - **Economic leadership**
  - **Academic leadership**

The Earth Sim… r in Japan

COMPUTENIK!

- Linpack benchm…
  TF/s = 87% of 40…
- Completed Apri…
- Driven by climat…
  earthquake simulation
- Gordon Bell Prize at SC2002





http://www.es.jamstec.go.jp/esrdc/eng/menu.html

Understanding and Prediction of Global Climate Change	Understanding of Plate Tectonics
Occurrence prediction of meteorological disaster	Understanding of long-range crustal movements
Occurrence prediction of El Niño	Understanding of mechanism of seismicity
Understanding of effect of global warming	Understanding of migration of underground water and materials transfer in strata
Establishment of simulation technology with 1km resolution	

Pragmatic U.S. Leadership is well established and unchallenged

- National Security – well established, proven, and successful programs
  - DOE/NNSA: ASCI
  - DOD : DOD Mod, NSA, DARPA
- Economic Leadership (see TOP500 List)
  - 9 of the TOP10 machines are US based and US built
  - 50% of the TOP500 machines are US based
  - More than 90% of the TOP500 machines are US built

U.S. Computational Science Leadership

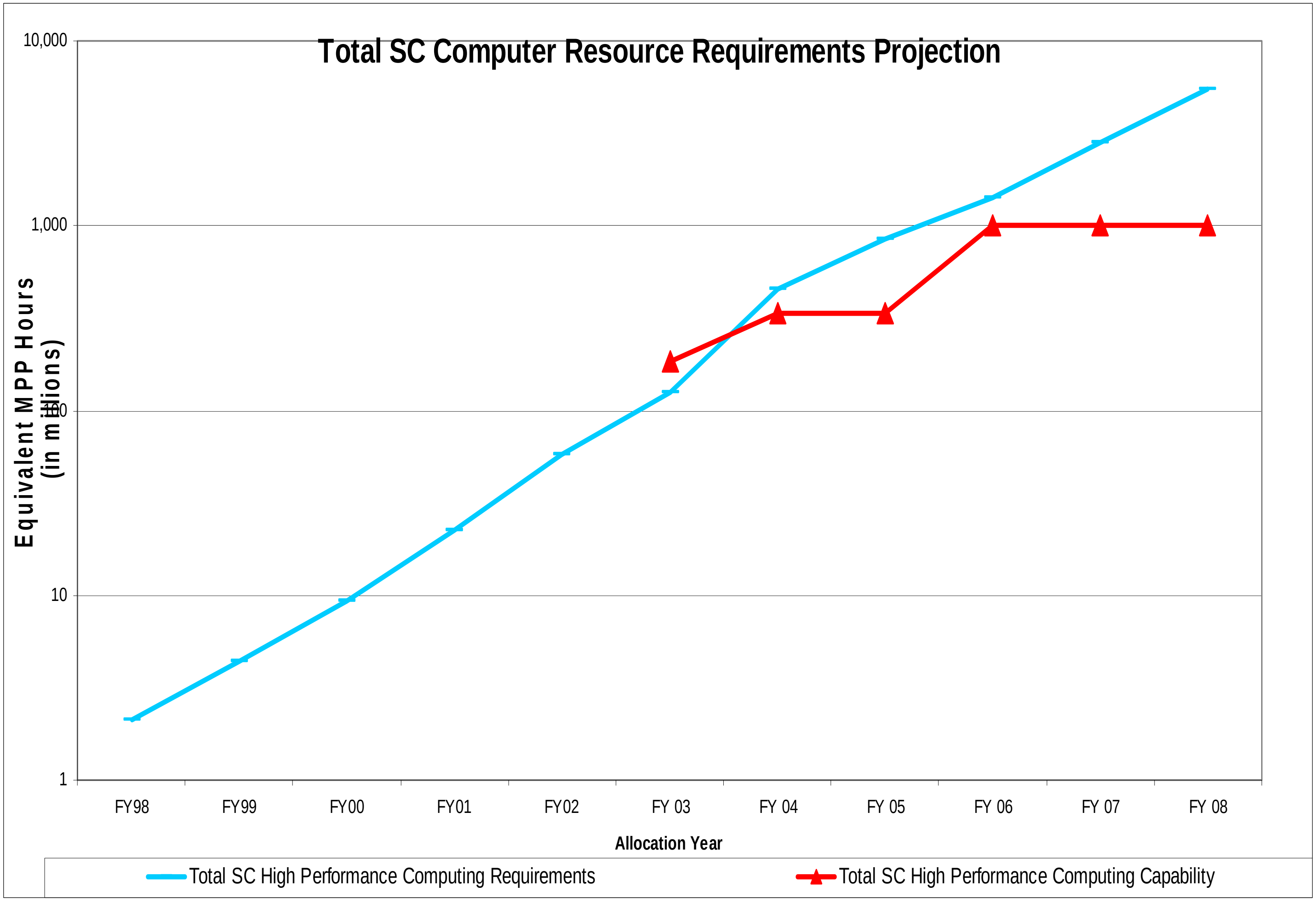
There is a large national investment in scientific software that is dedicated to current massively parallel hardware architectures

- **Scientific Discovery Through Advanced Computing (SciDAC) initiative in DOE**
- **Accelerated Strategic Computing Initiative (ASCI) in DOE**
- **Supercomputing Centers of the National Science Foundation (NCSA, NPACI, Pittsburgh)**
- **Cluster computing in universities and labs**

This is a strong a vibrant field. Computational Simulation is well established in the US as the third “leg” of science

- **See “Science Case for Large Scale Simulation” (SCaLeS), June 2003 www.pnl.gov/scales/**

Lack of Investment for HEC in Basic Sciences



Adapted from Dan Hitchcock, DOE -SC

Scientific Advances for 100 to 1000 times today's capability

Research Programs	Major Scientific Advances
Biological and Environmental Sciences	• **Provide global forecasts of Earth's future climate at regional scales using high-resolution, fully coupled and interactive climate, chemistry, land cover, and carbon cycle models.**
Chemical and Materials Sciences	• **Provide direct 3-dimensional simulations of a turbulent methane-air jet flame with detailed chemistry and direct 3-dimensional simulations of autoignition of *n*-heptane at high pressure, leading to more-efficient, lower-emission combustion devices.**
Fusion Energy Sciences	• **Improve understanding of fundamental physical phenomena in high-temperature plasmas, including transport of energy and particles, turbulence, global equilibrium and stability, magnetic reconnection, electromagnetic wave/particle interactions, boundary layer effects in plasmas, and plasma/material interactions.**
High Energy Physics	• **Establish the limits of the Standard Model of Elementary Particle Physics by achieving a detailed understanding of the effects of strong nuclear interactions in many different processes, so that the equality of Standard Model parameters measured in different experiments can be verified. (Or, if verification fails, signal the discovery of new physical phenomena at extreme sub-nuclear distances.)**

Loss of scientific leadership in computational science without a leadership class HEC platform

www.pnl.gov/scales

Outline

- Introducing NERSC
- The Evolution of HEC Systems and Applications: the Divergence Problem
- U.S. Leadership in HEC
- The Implementation of a Cyberinfrastructure

Rogers and Bozeman, pg. 30

- “ … tying the cyber infrastructure to one arbitrarily selected approach creates a distorting effect on the overall environment for the development of HEC technologies …”

- **Grids and clusters are highly important elements of our cyber infrastructure**

- **But they are NOT the one and only answer to many HEC requirements for basic science**

Current Trends in Computer Science Research in the US

Because of the emphasis on cyberinfrastructure the attention of research in computer science is not directed towards scientific supercomputing

- Primary focus is on Grids and Information Technology
- Only a handful of supercomputing relevant computer architecture projects currently exist at US universities; versus of the order of 50 in 1992
- Parallel language and tools research has been almost abandoned
- Petaflops Initiative (~1997) was not extended beyond the pilot study by any federal sponsors